

<https://helda.helsinki.fi>

Kurt Gödel's Anticipation of the Turing Machine : A Vitalistic Approach

Lethen, Tim

2020-07-02

Lethen , T 2020 , ' Kurt Gödel's Anticipation of the Turing Machine : A Vitalistic Approach ' ,
History and Philosophy of Logic , vol. 41 , no. 3 , pp. 252-264 . <https://doi.org/10.1080/01445340.2020.1768461>

<http://hdl.handle.net/10138/331641>

<https://doi.org/10.1080/01445340.2020.1768461>

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Kurt Gödel's anticipation of the Turing machine: a vitalistic approach

Tim Lethen*

Department of Philosophy, P.O. box 24 (Unioninkatu 40 A), FI-00014 University of Helsinki, Finland

ARTICLE HISTORY

Compiled June 29, 2020

ABSTRACT

In 1935/36 Kurt Gödel wrote three notebooks on the foundations of quantum mechanics, which have now been entirely transcribed for the first time. Whereas a lot of the material is rather technical in character, many of Gödel's remarks have a philosophical background and concentrate on Leibnizian monadology as well as on vitalism. Obviously influenced by the vitalistic writings of Hans Driesch and his 'proofs' for the existence of an entelechy in every living organism, Gödel briefly develops the idea of a computing machine which closely resembles Turing's groundbreaking conception. After introducing the notebooks on quantum mechanics, this article describes Gödel's vitalistic *Weltbild* and the ideas leading to the development of his computing machine. It investigates a notion of *lawlike sequence* which closely resembles Turing's concept of a *computable number* and which Gödel himself calls 'problematic', and compares it to the opposed concept of randomness, drawing upon the notion of program size complexity. Finally, Gödel's machine is implemented in a dialect of the LISP programming language.

KEYWORDS

Kurt Gödel; Turing machines; vitalism

AMS CLASSIFICATION

01A60; 03D10; 68Q05

1. Introduction

It is a well known fact that Kurt Gödel, during his whole adult life, devoted a great part of his time and energy to philosophy. According to Hao Wang (*Wang 1987*, p. 27), *all* of Gödel's important work 'is closely related to philosophy: in philosophy or philosophically motivated or having philosophical consequences or using philosophy as heuristic principles. It may even be supposed that, for him, all fundamental theoretical work is related to philosophy in one or more of the four nonexclusive ways.' This article will concentrate on the second way, philosophically motivated work, exemplarily pointing out an early anticipation of what is nowadays called the *Turing machine*, explaining how Gödel's thoughts about this topic may have been influenced by his early vitalistic views.

*Email: tim.lethen@helsinki.fi

Gödel's vitalistic tendencies are clearly expressed in a set of notebooks on quantum mechanics written in Vienna in 1935/36. In these books, which have now been entirely transcribed from the Gabelsberger shorthand system by the present author, Gödel, opposing materialism, not only carefully describes his vitalistic *Weltbild* but also briefly develops the idea of a computing machine, obviously drawing upon ideas taken from the writings of *Hans Driesch*, who may be described as the driving force of *neovitalism* in the early 20th century.

As a starting point, the article is briefly going to describe the notebooks on quantum mechanics and their contents. It will then introduce vitalism as part of Gödel's *Weltanschauungen* and describe one of Driesch's main ideas for a proof of the existence of an entelechy in every living organism.

Gödel's idea of a computing machine along with its basic properties will then be presented, emphasising the influence of an underlying philosophy. At the same time, connections are drawn to the Turing machine, its capabilities, and its information-theoretic restrictions.

Finally, an implementation of the machine as described in his notebooks on quantum mechanics is presented. It is written in a dialect of the programming language LISP which supports an easy implementation of a universal machine which tries to run binary programs up to a given time limit.

While usually most of the material concerning Gödel's *Weltanschauung* is taken either from works published during his lifetime, or from sources published posthumously, this article will take the opportunity and present a collection of unpublished material in connection with Gödel's philosophical views. These excerpts are mainly taken from notebooks in his *Nachlass* such as the *MaxPhil* series, and his notebooks on quantum mechanics, written in the mid 1930s and preserved in his *Nachlass* in Princeton. Transcriptions from the Gabelsberger shorthand system were produced by the author as part of the GÖDELIANA project based in Helsinki, Finland. The original German source material (linked by a mark $\langle n \rangle$ in the running text) is given in an appendix.

2. Gödel's notebooks on quantum mechanics

In 1935, Gödel wrote two complete notebooks¹, altogether nearly 200 pages, expressing his ideas, thoughts, and questions about the foundations of quantum mechanics, carefully devised into a systematic list of about 340 items. Whereas the earlier remarks cover mainly technical material, they become more and more philosophical, finally leading to what might be described as Gödel's first steps in philosophy, an attempt to apply Leibnizian monadology to quantum mechanics and physics in general, covering topics like free will, the direction of time, and the borderline between observer and experiment. The books also contain the compilation of what Gödel later calls his physical 'Arbeitsprogramm [working program]', a list of topics like 'Spin' and 'Proj. Rel. Theorie' he obviously planned to work on.

Only one year later, in a period of great personal instability, Gödel revised his own notes in a third book titled '*Aflenz 1936 (Analysis, Physik)*'.² While leaving out or simply just copying some of the earlier items, others are carefully reformulated

¹The books are simply titled '*Quantenmech. I*' and '*Quantenmech. II*'. Kurt Gödel Papers, Box 6b, Folder 78, item accessions 030106 and 030107, on deposit with the Manuscripts Division, Department of Rare Books and Special Collections, Princeton University Library. Used with permission of Institute for Advanced Study. Unpublished Copyright Institute for Advanced Study. All rights reserved.

²Kurt Gödel Papers, Box 6a, Folder 59, item accession 030082.

and extended, thus leading to an updated list of 143 items³. This revision offers the opportunity to follow the development of his ideas and thoughts and to judge upon the role they play in Gödel's considerations.

3. Neovitalism

In a draft of a lecture called 'The modern development of the foundations of mathematics in the light of philosophy' (*Gödel 1961*) written in 1961 or shortly thereafter, Gödel presents a general scheme of *Weltanschauungen*, ordered by their affinity to metaphysics or religion. Whereas scepticism, materialism and positivism stand on one side, spiritualism, idealism, and theology stand on the other. He observes a clear shift from 'right to left', i.e. a turning away from theology towards materialism, having reached its peak especially in the field of physics and leading to what he considers 'the end of all theoretical science.'

In the same draft, materialism is described as 'regard[ing] the world as an unordered and therefore meaningless heap of atoms.' It is well known that Gödel clearly opposed this view, and item 252 of *Quantenmechanik II*—strongly drawing upon Leibnizian monadistic views—clearly underlines this opposition:

The monad of an animal body is an existence, on the other hand somehow identical with organization [order of the parts]. This becomes understandable by photochemistry. As a result, an apparent substance (light quantum) transforms into an organization (complex structure). From the principle of conservation of individuals it follows that this complex structure is a certain simple substance. (1)

And a remark in his philosophical notebook *MaxPhil V*⁴ hints at the elements that materialism and positivism are lacking in order to suit Gödel's personal views. On page 344 he writes:

Remark (Philosophy): The materialistic (positivistic) worldview means to reality that there are only laws of pressure and shock [and otherwise only chaos] and only material things. Another form, that there are only sensations and laws of their succession. An extension is possible in two ways:

- 1.) As for the existing things: soul, concepts, angels.
- 2.) As for the existing laws (i.e., general facts): justice, superstition, etc.

Positivism is better in that at least no restriction of the laws, but in truth there are no laws formable at all. [...] Positivism is the only form of materialism that can still exist in the present. (2)

His vitalistic viewpoint is clearly exposed and carefully described in item 283⁵ of the notebook *Quantenmechanik II*:

Apparently, there are superior organizational forms.

1. unorganized ether maybe = light
2. electrons and protons
3. nuclei
4. atoms
5. molecules
6. living beings

³The items in this book are first numbered consecutively from 1 to 72. From then on, the numbers used are the same as the corresponding ones in the earlier two books, jumping over left out items.

⁴Kurt Gödel Papers, Box 6b, Folder 67, item accession 030091. This notebook was written between May and July 1942.

⁵The number 283 itself is actually missing but can be reconstructed from the numbering in the *Aflenz* book.

Although every thing of level n contains things of level $n - 1$ as part, it is not to be understood as an aggregate of these. That is, its behavior can not be explained by spatiotemporal laws regarding things of the $(n - 1)$ -th type. That is, the state (relation to each other) [necessary for the prediction of the behavior] of the elementary parts is not a spatial relationship. For example, two hydrogen atoms can stand in the relation of ‘forming a hydrogen molecule’ or in the relation of ‘forming two separate hydrogen atoms’ and then behave differently, though the probabilities of the spatiotemporal positions may be similar. Likewise, an atomic system may perhaps stand in the relationship of ‘forming an organism’ or ‘forming a cluster of atoms’, and behave differently accordingly. That is, an organism is not described by stating the space-time position of the atoms forming it, but it is also necessary to add something of the kind ‘and they form an organism’. Or there is also an entelechy. [This would be the difference between organism and machine, but would not exclude the possibility of an artificial production of life.] From this would follow, for example: Organization can not be explained by field-like effects propagated in the living substance [because these would have to be the same in the case of matter in an unorganized state]. (3)

This passage strongly resembles statements formulated by Hans Driesch, who can be regarded as the driving force of *Neovitalism* at the time (see *Driesch 1905, 1928*), and whose name is explicitly given in item 263 of *Quantenmechanik II*.⁶ In his Gifford lectures, published in 1908, Driesch writes (*Driesch 1928*, p. 15):

The organism is a specific individual body composed of a typical combination of different specific parts, each performing a particular physiological function. [...] but the organism is not an aggregate, not even for the most superficial consideration. (4)

And he continues (p. 36):

The adult organism is evidently a multi-level structure: it consists of organs as a partial whole, these of tissues, these of cells, these of intracellular substances, these of molecules, these of electrons. But we can now say of the development of the organism that, on its way from the egg to the adult, it transfers the living being *from a low-grade to a high-grade manifold*, at least as far as the visible is concerned. (5)

The connection between vitalism and quantum mechanics was certainly no stranger at the time. In his *Aftenz* book, Gödel writes:

Quantum mechanics favours the mystical solutions of the questions vitalism, second sight, *Gestalt* theory, relation between brain and consciousness, because the other solutions follow only from the assumption of spacetime physics as framework. The question ‘How does the process proceed spatiotemporally,’ which is the most fruitful means of analyzing a phenomenon, has for the first time proved infertile and meaningless in the process of radiation. Perhaps it is the same with the unexplained appearances in quotation marks above. (They are only inexplicable by asking this question, that is, by the use of space-time physics.) (6)

The phenomena in quotation marks Gödel is referring to include (amongst others) stigmatization⁷, the inheritance of acquired traits⁸, and the increasing of the number of twin births after the war⁹. It is worth noting how Driesch himself, of course without alluding to quantum mechanics, formulates the essence of the doctrine of the autonomy of life. In *Driesch 1928*, p. 126, he writes:

⁶Here Gödel writes: ‘Driesch, Entwicklung eines ganzen Organs aus einer halben Knospe.’ Other names mentioned in this remark are those of the biologists *Bleuler*, *Spemann*, and *Hartmann*.

⁷Gödel obviously was very interested in the stigmatization of *Therese von Konnersreuth*. Her name often appears in his theological notes.

⁸In a footnote to item 263 of *Quantenmechanik II*, Gödel refers to an article by Eugen Bleuler on this topic.

⁹In this connection, Gödel refers to ‘Arch. f. soz. u. Demogr. I, 1926, H.4’ in item 305 of *Quantenmechanik II*.

Certain events on living bodies are of such a nature that they can not be deduced from a knowledge of the coordinates (positions), forces, and velocities of the individual physical elements. ⟨7⟩

In order to complete the picture, we finish this section with two remarks about vitalism and entelechies which Gödel states in his philosophical notebooks *MaxPhil XI*¹⁰ (pp. 43) and *MaxPhil XIV*¹¹ (pp. 76), respectively. Both passages leave no doubt that Gödel, nearly ten years after writing his comments on the foundations of quantum mechanics, still abided by his vitalistic views. The first one was written in 1944 and reads:

Is the difference between the living and the lifeless maybe that its causal laws can not be captured in ‘mechanical’ rules, that is, not ‘formalized’? So that means a higher level of complication. The view contrary to vitalism simply asserts: all living things are dead. ⟨8⟩

The second remark was probably written by the end of the 1940s and emphasises an interesting role of the entelechy.

Separation is disharmony, union is harmony. So what we want in the end is to return to unity. [But maybe the evil creatures just want the separation?] So that means ‘balancing’ [equalization after Goldstein]. The same principle is also expressed in the laws of physics, and therefore the equation $\dot{V} = k\Delta V$ is so common. But this physical equalization principle is probably not the only one, and so there are entelechies, which are nothing but higher ‘harmonizers’. Perhaps even the particle problem and quantum mechanics require such ‘higher’ harmonizers [that is, in addition to the equalization expressed in the field equations]. ⟨9⟩

4. The Gödel machine

Alan M. Turing introduced his computing device, which has ever since been known as the *Turing machine*, in his groundbreaking paper ‘*On Computable Numbers, with an Application to the Entscheidungsproblem*’ (Turing 1936–7).¹² It is truly remarkable that Turing not only introduced the concept of a programmable machine, he also defined the notion of a universal machine, taking both its program and the required data on the same input device, a bidirectional infinite tape. At the same time he was immediately able to foresee the machines’ theoretical limitations, instantiated by the now famous *halting problem*. It is worth noting that during the same period¹³ Emil L. Post independently introduced a very similar concept in his paper ‘*Finite combinatory Processes, Formulation I*’ (Post 1936), one of the main differences being that he did not speak of a machine but rather of a ‘problem solver or worker’ working in a symbol space, again consisting of a ‘two way infinite sequence of spaces or boxes.’

Following Hao Wang (Wang 1987, p. 111), Gödel considered ‘Turing’s justification of the adequacy of his precise concept to the intuitive notion of computability as a piece of philosophical work.’ And in connection with this intuitive notion, Wang continues (p. 170): ‘It is, however, according to G., only Turing’s work of 1936 on computable numbers that for the first time presented a convincing analysis to show us the correct perspective by which to see the intuitive concept clearly’ and (p. 109) ‘He was probably

¹⁰Kurt Gödel Papers, Box 6b, Folder 70, item accession 030097. I am indebted to Maria Hämeen-Anttila for bringing this passage to my attention.

¹¹Kurt Gödel Papers, Box 6b, Folder 72, item accession 030099.

¹²The paper was received in May 1936 and published in January 1937.

¹³The paper was received in October 1936 and published in September 1936.

$a, 0 \rightarrow L, a$	
$a, 1 \rightarrow 1', L, a$	
$a, \begin{smallmatrix} \neq 0, 1 \\ x \end{smallmatrix} \rightarrow b$	a, b, c, d sind die ‘Stellungen’,
$b, \begin{smallmatrix} \neq 0 \\ x \end{smallmatrix} \rightarrow R, b$	$L = \text{links}, R = \text{rechts},$
$b, 0 \rightarrow \text{stop}$	$*, \bar{*}, 1, 1'', 1'''$ sind die ‘Zeichen.’

Figure 1.: Gödel’s Turing machine, replacing every 1 by a 1’. As usual, L and R denote a shift of the I/O-head, a and b are the machine’s possible states (or ‘Stellungen,’ as Gödel calls them).

surprised by Turing’s solution, which was more elegant and conclusive than he had expected.’¹⁴

Very soon, Gödel himself began programming Turing machines. On a single page found in his notebook *Arbeitsheft 6*¹⁵ and titled ‘Turing’sche Berechenbarkeit \equiv Herbrand’scher, Einschub’, he implements two small (slightly erroneous) programs, headed respectively:

- (1) Es gibt eine Maschine für die Veränderung aller 1 in 1’.
[There is a machine for changing every 1 into a 1’.]
- (2) Es gibt eine Maschine, welche ebenso viele 1’ aufschreibt, als 1 dastehen, nach einem * Zeichen.
[There is a machine which writes as many 1’ as there are 1, behind a * sign.]

If only for historical reasons, the first of these two programs is reproduced in Figure 1. The page ends with the comment ‘Ende Einschub’, and so far no connecting pages could be traced.¹⁶

Returning to the topic of vitalism, it is now of central importance that all of Driesch’s ‘proofs’ for the existence of an entelechy are based on the idea that certain phenomena within a living organisms simply cannot be reduced to a purely mechanistic behaviour of a machine or a collective of machines. Of course, in this connection the term *machine* has to be understood in the widest sense, and in *Driesch 1928* (p. 117), Driesch himself describes it as follows:

We will understand the word ‘machine’ in its most general sense. *For us, a machine is therefore a typical arrangement of physical and chemical constituents, through the impact of which a typical effect is achieved.* (10)

The proof which might have influenced Gödel most is described by Driesch in *Driesch 1905* as his ‘second proof, based on the genesis of complex equipotential systems.’ His experiments with the egg of the sea urchin, in which he agitated the

¹⁴In his Gibbs Lecture *Gödel 1951*, Gödel expresses his opinion as follows: ‘The most satisfactory way, in my opinion, is that of reducing the concept of finite procedure to that of a machine with a finite number of parts, as has been done by the British mathematician Turing.’

¹⁵Kurt Gödel Papers, Box 5c, Folder 18, item accession 030024.

¹⁶I am indebted to Jan von Plato for pointing out that Gödel briefly refers to this page in a notebook called *Excerptenheft 7* (Kurt Gödel Papers, Box 6a, Folder 56, item accession 030080). Here, in a list of eight items probably written in 1945 and simply headed ‘Lesen’, the last item reads ‘Idee Turing A.H.6 p.-4.’

embryo at the two-cell and four-cell stage until it disassociated into separate cells each of which then developed normally, led him to the conclusion that there should be some complexity inherent in the very first cell which is completely preserved when this cell splits into two. Driesch then argues that indeed no machine can keep its complex structure when it is divided into two or more parts.

Also, as Driesch repeatedly notes, the cells of the embryo include a *prospective potency* or *prospective power* which enables them not only to develop parts which they would have developed if the disturbing experiments would not have been conducted, but also much more specialized parts of the growing organism.

This kind of reasoning clearly leads to the conclusion that a (without doubt highly complex) hypothetical machine within the egg ('Keim') would have to be capable of reproducing itself as well as producing every specialized machine necessary for a complete development of the striven mature organism.

Many items in Gödel's books on quantum mechanics leave no doubt that he must have been very familiar with this kind of neovitalistic reasoning based on the impossibility of the existence of a machine. And indeed, a short time before Turing and Post had published their ideas on computability, Gödel had already written down the first hints concerning his thoughts about his own (vitalistic) machine in 1935. In *Quantenmech. II* we read the following note:

317. The machine that constructs a diagonal sequence of lawlike machines would be one that produces more and more organization from unorganized matter, hence something akin to a 'germ' ('Keim'). (11)

And in his *Aflenz* book, written about one year later, we can see the following reformulation:

320. Problem: Definition of lawlike sequence as: sequence producible by a machine [Machine that produces a sequence of machines, and this diagonal sequence would be something that transforms more and more unorganized matter into organized matter (germ).] (12)

The association to organized and unorganized matter as well as the clear allusion to the 'Keim' leave no doubt as to where the idea of a computing machine originates from. At the same time it should be noted that the concept of '*Diagonalfolge*' is of course the actual essence both of Gödel's incompleteness proof and of Turing's main results concerning the halting problem.

It is now possible to deduce certain formal properties of what we are going to call the *Gödel machine*.

- The machine has an output device to which it can write (possibly infinite) sequences of data.
- The type of the elements of the output sequences includes the set of *machines* or at least of some suitable encodings of these *machines*.
- As the output of a machine again can be machines, we can identify the terms *machine* and *machine description*.
- The fact that Gödel speaks of a '*Diagonalfolge*' seems to imply that any proposed set of possible machines is enumerable and can be diagonalized.

It remains an open question if the machines Gödel had in mind were also supposed to be equipped with an input device. In what follows, the machines are simply assumed to start their operations 'on the empty tape'. Note that the term *Gödel machine* will be used both for the general concept and for specific machines instantiating this very concept.

5. Lawlike sequences

The ‘problem’ Gödel encounters in item 320 of his *Aflenz* notebook is based on an ambiguity of the term *lawlike sequence* [*gesetzmäßige Folge*]. In this section, two conflicting definitions of this term are given, and each definition is applied to the *diagonal sequence* [*Diagonalfolge*] mentioned by Gödel. To this end, we fix an enumeration m_0, m_1, \dots of all possible Gödel machines and denote the (infinite) output sequence of machine m_i by $(s_i^j)_{j \in \mathbb{N}}$. If a machine m_i has only finitely many outputs (either because it halts or gets stuck in an infinite loop without any further outputs) and the last output is s_i^k , we let $s_i^l = 0$ for each $l > k$. Without loss of generality we assume $s_i^j \in \mathbb{N}$ for $i, j \in \mathbb{N}$. The sequence $(s_i^i)_{i \in \mathbb{N}}$ is then called a *diagonal sequence* [*Diagonalfolge*].

One can now formalize the ‘problematic’ definition taken from item 320 as follows.

Definition 1: An infinite sequence of natural numbers $(t_j)_{j \in \mathbb{N}}$ is *lawlike* [*gesetzmäßig*] iff there exists an $i \in \mathbb{N}$ with $t_j = s_i^j$ for all $j \in \mathbb{N}$. In other words, there has to be a Gödel machine with the output sequence $(t_j)_{j \in \mathbb{N}}$.

It is now immediately clear that, following this definition, the diagonal sequence is not lawlike, because—due to the usual conflict on the diagonal—there can be no machine on the list producing the closely related sequence $(s_i^i + 1)_{i \in \mathbb{N}}$.¹⁷ Intuitively, this result feels ‘problematic’, because the diagonal sequence follows a very simple *building law*. It should be very interesting to note how closely Gödel’s ideas resemble those of Turing, if only the term *lawlike sequence* be changed to *computable sequence*.¹⁸

At this point it seems to be useful to introduce the notion of *randomness* as a natural counterpart to *lawlikeness*.¹⁹ The second definition will therefore define randomness of infinite sequences of natural numbers, taking up ideas from the field of *Algorithmic Information Theory* (see *Chaitin 1987; Calude 2002*). For this purpose, we will first adopt and customise the definition of a (universal) *computer*. This in turn will lead to the notion of *program size complexity* which will enable us to capture the meaning of *randomness* in connection with infinite sequences of natural numbers. In the definitions, A is a finite alphabet. A^* denotes the set of finite words over the alphabet A .

Definition 2.1: A *Chaitin computer* is a partial computable function $C : A^* \rightarrow A^*$ such that the domain of C is prefix-free.

Note that the computers defined here take only programs as their input, i.e. the programs themselves do not operate on any data.²⁰ The prefix-free domain ensures that programs are *self-delimiting*: a computer knows where its programs end.²¹

Definition 2.2: A Chaitin computer C is *universal* if for each Chaitin computer D there is a constant c with the following property: if $C(x)$ is defined, then there exists

¹⁷Here it is assumed that the lawlikeness of the sequence $(s_i^i)_{i \in \mathbb{N}}$ implies the lawlikeness of the sequence $(s_i^i + 1)_{i \in \mathbb{N}}$.

¹⁸In *Turing 1936–7*, Turing stresses the interchangeability between the terms *computable number* and *computable sequence*.

¹⁹Although there seems to be no direct influence, it is interesting to note that Driesch explicitly identifies the terms ‘random’ (‘zufällig’) and ‘lawless’ (‘gesetlos’) in *Driesch 1928*, p. 374.

²⁰Usually, the domain of a computer is defined as $A^* \times A^*$, which is interpreted as the set of program/data pairs. While the data plays a crucial role in the definition of *conditional complexity*, it is set to the empty word if (as in the present situation) only unconditional complexities are considered. See *Chaitin 1987, Calude 2002* for further details.

²¹The computers operating on self-delimiting programs are named after *Gregory Chaitin* who first considered this property.

a string x' such that $D(x') = C(x)$ and $|x'| \leq |x| + c$, where $|x|$ denotes the length of the string x .

A universal Chaitin computer is thus able to imitate any other Chaitin computer D by simply prefixing to the programs of D an explanation (of length c) how D works. We can now fix a universal Chaitin computer U and define the *program-size complexity* H of an output string x as the length of the shortest program for U computing x .

Definition 2.3: The *program-size complexity* is the partial function $H : A^* \rightarrow \mathbb{N}$ with $H(x) = \min \{|p| : p \in A^*, U(p) = x\}$.

Whereas the program-size complexity does indeed depend on the fixed universal computer U , it can be shown that the corresponding complexities differ only at most by a constant value c (see *Calude 2002*). Finally, we can now define *randomness* of infinite strings. The basic idea is to consider the set of finite prefixes of this infinite string, stipulating the algorithmical incompressibility of these prefixes. If A^ω denotes the set of infinite sequences over the alphabet A and $\mathbf{x} \in A^\omega$, $\mathbf{x}(n)$ denotes the finite prefix of \mathbf{x} of length n .

Definition 2.4: A sequence $\mathbf{x} \in A^\omega$ is *random* iff there exists a natural c such that $H(\mathbf{x}(n)) \geq n - c$ for all natural $n \geq 1$. A sequence $\mathbf{x} \in A^\omega$ is *lawlike* [*gesetzmäßig*] iff it is not random.

In other words, an infinite sequence is lawlike iff infinitely many of its prefixes can be algorithmically compressed, which is now indeed possible with Gödel's diagonal sequence. To demonstrate this, we consider a prefix of the diagonal of length n . A program computing this prefix may now proceed as follows: if the number n and the number k of the first n Gödel machines that reach the diagonal are hardwired into the program, it constructs and simulates these n machines until k of them have reached the diagonal. The remaining diagonal elements are then set to zero. Note that such a program would consist of $2 \log_2 n + c$ bits and therefore, for large enough n , is indeed much shorter than the prefix itself.

6. An Implementation of Gödel's Keim-machine

Gödel's machine as introduced in his notebooks on quantum mechanics is remarkable for the fact that it obviously seems to enumerate all possible machine descriptions and at the same time simulates their actual behaviour, at least up to the diagonal element. Machines of this type have indeed been considered in the literature in more recent times, notably by Gregory Chaitin who introduces a machine calculating the halting probability Ω by enumerating and running all possible machine descriptions (*Chaitin 1998*). Chaitin uses a SCHEME-like dialect of the LISP programming language, fixing a very limited number of primitive functions mainly for list manipulation and arithmetic operations. The main new ingredient which makes his dialect very suitable for programming in an algorithmic information theory context, is the function `try`. It takes three arguments: (1) a natural number limiting the number of execution steps (realized by a limited depth of the recursive descend of the executing interpreter), (2) a LISP expression which is tried to be evaluated, and (3) a list of binary data used as an environment accessible by the evaluated expression through the built-in functions `read-bit` and `read-exp`, the latter reading a complete LISP expression. `try` always returns a list of three values: an indicator if the evaluation was successful (`success/failure`), the value of the expression or the reason for not being success-

ful (out-of-time/out-of-data), and a (chronological) list of caught outputs of the evaluated expression, thrown by the function `display`.²²

It is exactly the function `try` that enables us to actually implement and run Gödel's *Keim machine*, given in Figure 2, in an amazingly compact way. Gödel machines are taken to be binary strings representing a piece of software consisting of a LISP prefix possibly followed by further binary data used by the prefix itself. The Keim machine maintains a list of these binary programs. In each step of an infinite loop, the longest programs are extended by the bits 1 and 0. (To this end, the *prog-list* contains a list of the longest programs as its `car`, whereas the complete list of programs can be found as its `cdr`.) After having extended the *prog-list*, each program is executed up to a certain time limit which increases while the *prog-list* is processed.

When a single binary program (with number n) is executed within a certain time limit, several cases may occur.

- (1) The program has already produced at least n outputs. In this case the diagonal element can be displayed as an output of the Keim machine.
- (2) Otherwise we consider two subcases:
 - (a) The execution of the program has been stopped running out of time. In this case no output is displayed. (The program will be executed again with a larger time limit in the next round.)
 - (b) If the program has terminated regularly or because it ran out of data, a 0 is displayed as its diagonal element.

Note that no program is ever deleted from the *prog-list* and that therefore the same outputs are produced again and again. In order to identify these outputs, they are combined with the number of the producing program. This phenomenon might raise the issue of efficiency of the Keim machine. However, as there are infinitely many *busy-beaver*-like programs on the list whose number of steps grows faster than any computable function, it is not worth following this issue.

As the Keim machine presented here is itself a Gödel machine, it will eventually produce and execute itself. Consisting of 10784 bits²³ in its binary version, it will eventually be found in the *prog-list* among the first $2^{10785} - 1$ programs.

7. Conclusion

The discovery of Gödel's notes about a computing machine certainly does not lead to the necessity of rewriting the history of the invention of the Turing machine. However, it does offer yet another jigsaw piece of Gödel's rich contribution to the modern information based society.

It is interesting to note that, while Francis Crick has been reported in *Chaitin 2012* as having been influenced by John von Neumann's ideas about self-reproducing automata, we now encounter an example of the opposite direction: Hans Driesch's biological experiments with the embryos of the sea urchin led him to his 'philosophy of the organism' which in turn inspired Gödel to reflect about computing machines and the problems arising 'on the diagonal.'

²²For an introduction to Chaitin's LISP dialect see *Chaitin 1998* or *Chaitin 2001*. *Chaitin 1998* also provides an interpreter written in the MATHEMATICA language.

²³In order to count the bits, 'syntactic sugar' has to be removed from the program and for instance all `let`-bindings have to be converted into `lambda`-expressions. See *Chaitin 2001* for details.

```

define (keim prog-list)
  [Gödel's Keim-machine.]

[Program extension]

let (extend-prog prog)
  [Extends a single program by 1 and by 0.]
  cons append '(1) prog
  cons append '(0) prog
  nil

let (extend-progs progs)
  [Extends a list of programs, using extend-prog.]
  if = length progs 0
    nil
    append (extend-prog car progs)
            (extend-progs cdr progs)

let (extend-prog-list prog-list)
  [Extends a prog-list whose car are the longest programs
   and whose cdr are all programs.]
  let extension (extend-progs car prog-list)
  cons extension
    append extension
      cdr prog-list

[Program execution]

let (run-progs progs n limit)
  [Executes progs within a limited number of steps.]
  if > length progs 0
    let temp (run car progs n limit)
    (run-progs cdr progs
      - n 1
      + limit 1)
  nil

let (nth pos list)
  [Return the n-th element of the list.]
  if = pos 0
    car list
    (nth (- pos 1) cdr list)

let (run prog n limit)
  [Executes the n-th prog within a limited number of steps.
   If possible, displays the program along with its diagonal element.]
  let result try limit '(eval read-exp) prog
  let outputs caddr result
  if < n length outputs
    display cons n cons (nth n outputs) nil
  if = out-of-time cadr result
    nil
    display cons n cons 0 nil

let progs cdr prog-list
let temp (run-progs progs
  - length progs 1
  0)
(keim (extend-prog-list prog-list))

```

Figure 2.: Gödel's Keim machine implemented in LISP. The machine executes a list of binary programs, increasing the time limit while the programs are getting shorter. The list is then extended by a set of programs of increased length. This process is repeated infinitely and started by the expression `(keim '(((()) ()))`, the argument representing the list containing only one (empty) program.

Gödel himself often stressed the fruitfulness of the identification of analogies.²⁴ And although his extensive notes on quantum mechanics reveal that vitalism was part of his *Weltbild*, it should be stressed that, when drawing upon such analogies, there is certainly no need to actually accept their underlying ideas and concepts: Even the strongest opposition to a darwinistic theory of evolution can not deny the success of the so-called evolutionary algorithms which rely heavily on Darwin's ideas. The inspiring impact of Driesch's experiments on Gödel's thinking should be regarded in exactly the same way.

Acknowledgements and Funding

The research for this article is a part of the GÖDELIANA project led by Jan von Plato, to whom I remain grateful for his encouraging support. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 787758) and from the Academy of Finland (Decision No. 318066).

References

- Calude, C. S. 2002. *Information and Randomness: An Algorithmic Perspective*. 2nd ed., Berlin: Springer.
- Chaitin, G. 1987. *Algorithmic Information Theory*. Cambridge: Cambridge University Press.
- Chaitin, G. 1998. *The Limits of Mathematics: a Course on Information Theory and the Limits of Formal Reasoning*. London: Springer.
- Chaitin, G. 2001. *Exploring Randomness*. London: Springer.
- Chaitin, G. 2012. *Proving Darwin: making biology mathematical*. New York: Pantheon Books.
- Driesch, H. 1905. *Der Vitalismus als Geschichte und als Lehre*. Leipzig: Johann Ambrosius Barth.
- Driesch, H. 1928. *Philosophie des Organischen*. 4th ed., Leipzig: Quelle & Meyer.
- Gödel, K. 1951. 'Some basic theorems on the foundations of mathematics and their [philosophical]²⁵ implications.' Printed in *Gödel 1995*, 304–323.
- Gödel, K. 1961. 'The modern development of the foundations of mathematics in the light of philosophy.' Printed in *Gödel 1995*, 374–387.
- Gödel, K. 1995. *Collected Works. Vol. III. Unpublished Essays and Lectures*. Oxford: Oxford University Press.
- Post E. L. 1936, 'Finite combinatory processes—formulation 1.' *The Journal of Symbolic Logic* 1, 103–105
- Turing, A. M. 1936–7. 'On computable numbers, with an application to the Entscheidungsproblem.' *Proceedings of the London Mathematical Society* (2) 42 (1936–7), 230–265. A correction, *ibid.* 43 (1937), 544–546
- Wang, H. 1987. *Reflections on Kurt Gödel*. Cambridge, Mass.: The MIT Press.

²⁴For example, in a list of nearly a hundred theological questions and remarks, Gödel notes: 'Ein sehr fruchbarer Gesichtspunkt ist anscheinend die Analogie: Ideen–Engel. (Wie überhaupt jede Analogie.)'

²⁵The word 'philosophical' was inadvertently omitted from the title of Gödel's Gibbs Lecture by the editors of vol. III of his *Collected Works Gödel 1995* (as noted in the Corrigenda to vol. III on p. 637 of vol. IV).

Appendix A. Original German versions

1. Die Monade eines tierischen Körpers ist eine Existenz, andererseits irgendwie identisch mit Organisation [Ordnung der Teile]. Dies wird verständlich durch die *Photochemie*. Dadurch wandelt sich eine offenbare *Substanz* (Lichtquant) um in eine Organisation (komplexe Struktur). Aus dem Prinzip der Erhaltung der Individuen folgt daher, dass diese komplexe Struktur eine gewisse einfache *Subst.* ist.
2. Bem. (Phil.): Die mat. (posit.) Weltanschauung bedeutet auf die Wirklichkeit bezogen, dass es nur Gesetze von Druck und Stoß [und sonst nur das Chaos] gibt und nur materielle Dinge. Eine andere Form, dass es nur Empfindungen und Gesetze ihrer Aufeinanderfolge gibt. Eine Überschreitung (ist) in zweifacher Richtung möglich:
 - 1.) Was die existierenden Dinge betrifft: Seele, Begriffe, Engel.
 - 2.) Was die bestehenden Gesetzmäßigkeiten (d.h. allgemeinen Sachverhalte) betrifft: Gerechtigkeit, Aberglaube, etc.
 Posit. ist insofern besser, als wenigstens keine Einschränkung der Gesetze, aber in Wahrheit (sind) überhaupt keine Gesetze formulierbar. [...] Der Posit. ist die einzige Form des Mat., die in der Gegenwart noch bestehen kann.
3. Es gibt scheinbar übergeordnete Organisationsformen.
 1. unorganisierter Äther vielleicht = Licht
 2. Elektronen und Protonen
 3. Kerne
 4. Atome
 5. Moleküle
 6. Lebewesen
 Obwohl jedes Ding der Stufe n Dinge der Stufe $n - 1$ als Teil enthält, so ist es nicht als räumliches Aggregat dieser zu verstehen. D.h., sein Verhalten kann nicht durch raumzeitliche Gesetze bezüglich der Dinge des $(n - 1)$ -Typs erklärt werden. D.h., der [für die Voraussage des Verhaltens nötige] Zustand (Verhältnis zueinander) der Elementarteile ist kein räumliches Verhältnis. Z.B. zwei Wasserstoff (Atome) können stehen in dem Verhältnis des 'ein Wasserstoffmolekül bilden' oder in dem Verhältnis des 'zwei getrennte Wasserstoffatome bilden' und verhalten sich dann verschieden, obwohl die Wahrscheinlichkeiten der raumzeitlichen Lage vielleicht ähnlich sind. Ebenso kann ein Atomsystem vielleicht in dem Verhältnis des 'einen Organismus bilden' stehen oder in dem Verhältnis 'einen Atomhaufen bilden' und dementsprechend sich anders verhalten. D.h., ein Organismus wird nicht dadurch beschrieben, dass man die raumzeitliche Lage der ihn bildenden Atome angibt, sondern man muss außerdem etwas hinzufügen von der Art 'und sie bilden einen Organismus'. Oder es ist noch eine Entel. dabei. [Diese wäre der Unterschied zwischen Organismus und Maschine, würde aber nicht die Möglichkeit einer künstlichen Herstellung von Leben ausschließen.] Daraus würde folgen etwa: Organisation kann nicht erklärt werden durch feldmäßige, in der lebenden Subst. fortgepflanzte Wirkungen [denn diese müssten ja bei Materie im unorganisierten Zustand dieselben sein].
4. Der Organismus ist ein spezifischer individueller Körper, aufgebaut von einer typischen Kombination verschiedener spezifischer Teile, welche jeweils eine bestimmte physiologische Funktion ausüben. [...] aber der Organismus ist nicht ein Aggregat, nicht einmal für die oberflächlichste Betrachtung.
5. Der erwachsene Organismus ist nun offenbar ein *mehrstufiges* Gebilde: es besteht aus Organen als Teilganzen, diese aus Geweben, diese aus Zellen, diese aus Intrazellulärsubstanzen, diese aus Molekülen, diese aus Elektronen. Von der Entwicklung des Organismus aber können wir nunmehr sagen, dass sie auf ihrem Weg vom Ei zum Erwachsenen das Lebewesen *aus einer wenigstufigen in eine hochstufige Mannigfaltigkeit überführe*, wenigstens soweit Sichtbares in Frage kommt.
6. Die Quantenmechanik begünstigt die mystischen Lösungen der Fragen Vitalismus, Hellsehen, Gestalttheorie, Verhältnis von Gehirn und Bewußtsein, weil die anderen Lösungen lediglich aus der Annahme einer Raumzeit-Physik als Rahmen folgen.

Die Frage 'Wie verläuft der Vorgang im Einzelnen raumzeitlich', die das fruchtbarste Mittel der Analyse einer Erscheinung ist, hat sich beim Prozess der Ausstrahlung zum ersten Mal als unfruchtbar und sinnlos erwiesen. Vielleicht ist es bei den oben

- unter Anführungszeichen stehenden unerklärlichen Erscheinungen ebenso. (Sie sind nur unerklärlich durch Stellung dieser Frage, d.h. Zugrundelegung der Raumzeit-Physik.)
7. Gewisse Geschehnisse an belebten Körpern sind von einer solchen Art, daß sie sich nicht aus einer Kenntnis der Koordinaten (Lagen), Kräfte und Geschwindigkeiten der einzelnen körperlichen Elemente herleiten lassen.
 8. Ist der Unterschied des Lebendigen zum Leblosen vielleicht der, dass sich seine Wirkungsgesetze nicht in 'mechanische' Regeln fassen, d.h. nicht 'formalisieren' lassen? D.h. also eine höhere Stufe der *Komplikation*. Die dem Vitalismus entgegengesetzte Anschauung behauptet also einfach: Alles Lebendige ist tot.
 9. Die Trennung ist die *Disharm.*, die Vereinigung die Harmonie. Was wir letztenendes wollen, ist also zur Einheit zurückkehren. [Aber vielleicht wollen die bösen Geschöpfe gerade die Trennung?] D.h. also, ein 'Ausgleich' [*equalization* nach *Goldstein*]. Dasselbe Prinzip (ist) auch in den physikalischen Gesetzen ausgedrückt, und daher (ist) die Gleichung $\ddot{V} = k\Delta V$ so häufig. Dieses physikalische Equalisationsprinzip (ist) aber wahrscheinlich nicht das einzige, und insofern gibt es *Entelech.*, welche nichts anderes sind als höhere 'Harmonizers'. Vielleicht sind schon für das Teilchenproblem und die Quantenmechanik solche 'höheren' *Harmonizers* nötig [d.h. neben der in den Feldgleichungen ausgedrückten *equalization* liegend].
 10. Wir werden das Wort 'Maschine' in seinem allgemeinsten Sinne verstehen. *Eine Maschine ist uns also ein typische Anordnung physikalischer und chemischer Konstituenten, durch deren Wirkung ein typischer Effekt erreicht wird.*
 11. 317. Die Maschine, welche eine Diagonalfolge einer Folge gesetzmäßiger²⁶ gesetzmäßiger Maschinen konstruiert, wäre eine solche, welche immer mehr Organisation aus unorganisierter Materie hervorbringt, also etwas ähnliches wie ein 'Keim'.
 12. 320. Problem: Definition von gesetzmäßiger Folge als: durch Maschine herstellbare Folge [Maschine, welche eine Folge von Maschinen herstellt, und diese Diagonalfolge wäre etwas, was immer mehr unorganisierte Materie in organisierte verwandelt (Keim).]

²⁶Actually, the word 'einer' is not cancelled here. This would allow for the following reading: 'Die Maschine, welche eine Diagonalfolge einer gesetzmäßigen Maschine konstruiert, wäre eine solche, [...]' It is assumed that Gödel simply forgot to cross the word out.